Biochemical Society Transactions www.biochemsoctrans.org

Bringing Together Biomolecular Simulation and Experimental Studies

A Biochemical Society Focused Meeting in conjunction with the Molecular Graphics and Modelling Society held at Manchester Interdisciplinary Biocentre, Manchester, U.K., 10–11 September 2007. Organized and Edited by Mike Sutcliffe (Manchester, U.K.).

Biomolecular simulation: historical picture and future perspectives

Wilfred F. van Gunsteren*[1] and Jožica Dolenc*†

*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zurich, Switzerland, and †Faculty of Chemistry and Chemical Technology, University of Ljubljana, SI-1000 Ljubljana, Slovenia

Abstract

Over the last 30 years, computation based on molecular models is playing an increasingly important role in biology, biological chemistry and biophysics. Since only a very limited number of properties of biomolecular systems are actually accessible to measurement by experimental means, computer simulation complements experiments by providing not only averages, but also distributions and time series of any definable, observable or non-observable, quantity. Biomolecular simulation may be used (i) to interpret experimental data, (ii) to provoke new experiments, (iii) to replace experiments and (iv) to protect intellectual property. Progress over the last 30 years is sketched and perspectives are outlined for the future.

Introduction

Over the last 30 years, simulation of the motion of biomolecular systems at the atomic level based on molecular models has played an increasingly important role in biological chemistry and physics [1–8]. This role is, however, still limited, because the available computing power, although growing fast, sets limits to the size of the systems that can be simulated, the time scale that can be covered and the accuracy that can be reached. Notwithstanding these limitations, computer simulation of the behaviour of the biomolecular systems is being practised, because it complements the experimental methodology of investigation. Only a very limited number of properties of biomolecular systems are actually accessible to measurement by experimental means. Moreover, experiments generally provide averages, over space (molecules) or time, of measurable quantities, the distribution of such quantities, again over time or space, remaining inaccessible, at least for microscopic space and time scales. This is where computer simulation has its strength by providing not only averages, but also distributions and time series of any definable, observable or non-observable, quantity, for example, conformational distributions or interactions between separate parts of molecular systems. A second advantage of computer simulation of (bio)molecular systems is that it can be used to investigate cause–effect relationships by individually changing model parameters that cannot be changed without affecting other parameters experimentally. For example, the effect of changing the solvent viscosity (by changing the mass of solvent molecules) on the folding rate of a polypeptide has been investigated using MD (Molecular Dynamics) simulation [9] without changing any other solvent parameter or property. Experimentally, it is not possible to change single properties leaving all others untouched.

In this very brief report, we outline, in the next section, the major developments in the area of biomolecular simulation since the first biomolecular dynamics simulations published 30 years ago [1,10,11]. In the section 'Is biomolecular dynamics simulation useful?', we give examples taken from our own work illustrating the usefulness of MD simulations (i) to interpret experimental data, (ii) to provoke new experiments, (iii) to replace experiments and (iv) to protect intellectual property. In the final section, we outline perspectives for the future.

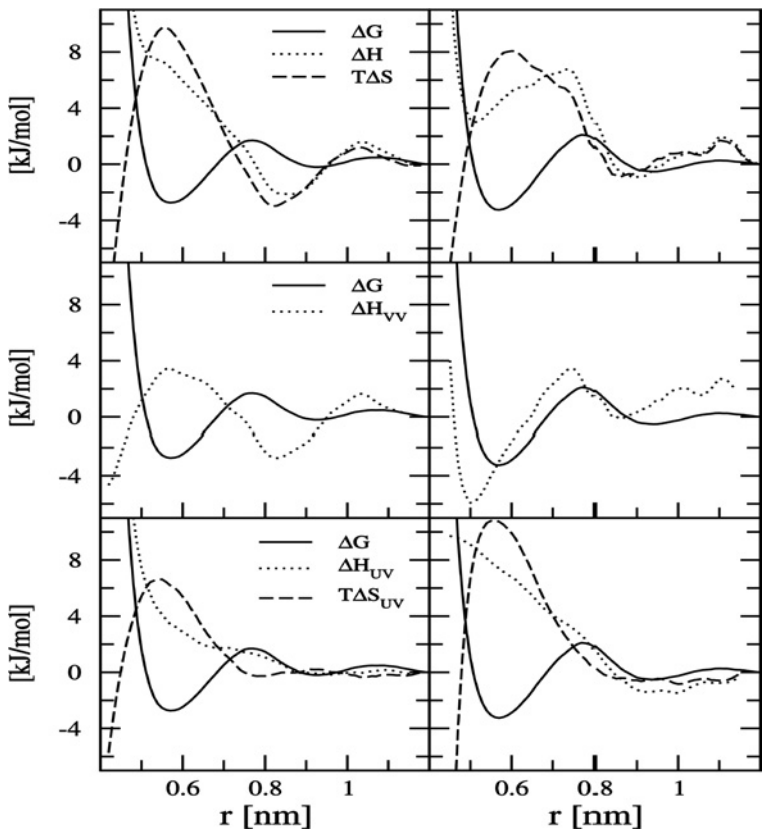
Thirty years of biomolecular dynamics simulation

A prerequisite for any molecular simulation is a molecular model that specifies the various types of interactions between

Key words: biomolecule, computer simulation, distribution, Molecular Dynamics, sampling.
Abbreviation used: MD, Molecular Dynamics.
[1]To whom correspondence should be addressed (email wfvgn@igc.phys.chem.ethz.ch).
Biochem. Soc. Trans. (2008) **36**, 11–15; doi:10.1042/BST0360011

Figure 1 | **Driving force for molecular association**

Free enthalpy of association (ΔG) and its enthalpic (ΔH) and entropic ($T\Delta S$) contributions (top panels), its solvent–solvent ($\Delta H_{VV} = T\Delta S_{VV}$) contributions (middle panels) and its solute–solvent (ΔH_{UV}, $T\Delta S_{UV}$) contributions (bottom panels) as a function of the distance between two *neo*-pentane molecules in pure water (left panels) and in urea solution (right panels) at 1 atm (1 atm ≡ 101.325 kPa) and 298 K.



the atoms in the systems. Such a model is often called a force field [12]. The first force fields for biomolecules were proposed by Lifson and Warshel [13]. Over the years, a number of general biomolecular force fields emerged and are still being further extended and refined [4,14–19]. As the word rightly conveys, in a molecular simulation, the motion along the (chosen) degrees of freedom is simulated by solving equations of motion, i.e. those of Newton, Hamilton or Lagrange, or more recently, when simulating quantum dynamics [20], those of Schrödinger.

MD simulation of proteins began in 1976 with an 8-week CECAM workshop on protein dynamics [1], which resulted, for example, in the publications mentioned above [10,11]. Since then, MD simulation of biomolecules has developed in different directions: (i) by expanding the number and types of degrees of freedom, (ii) by refinement of the molecular models, (iii) by improvement of the search and sampling power of MD algorithms and (iv) by expanding its range of applications in terms of types of molecules and processes simulated. Here, we only mention some examples.

The first MD simulations of proteins neglected solvent degrees of freedom [10,11]. Introduction of an explicit solvent in atomic [21] or molecular form [22] did make the simulations more realistic. Enzyme reactions could only be simulated by introducing quantum degrees of freedom [5,23,24]. Minimizing the effects of approximating long-range electrostatic interactions in one way or the other is still an issue [25–27]. Although introduction of atomic polarizability was already an issue at the CECAM workshop of 1976, only recently biomolecular force fields including this feature have been published [28–30]. During the 1980s, thermodynamic perturbation and integration techniques [31] were applied to calculate complexation free energies for proteins and other biomolecules. The calculation of the corresponding entropies is still in its infancy [32]. The computation of non-dynamic quantities did benefit a lot from the application of a variety of search and sampling enhancement techniques. For a recent review of these, see [33]. Sampling techniques, such as replica-exchange [34], which can exploit massively parallel computers [35], are currently widely used. Here, we can also mention the Folding@home project (http://folding.stanford.edu/) that exploits thousands of computers all over the world [36]. MD simulation is now routinely used to investigate the stability of proteins, to simulate the polypeptide folding

process and in ligand docking. It is a standard tool in protein structure refinement based on NMR [37] or X-ray [38] data.

Is biomolecular dynamics simulation useful?

Computer simulation trajectories can be used to interpret experimentally measured data, e.g. the difference in stability between protein mutants [39], or to resolve seeming contradictions between NMR and X-ray data on the same protein [40]. Figure 1 shows an example of the complementarity of simulation and experiments regarding the driving forces for molecular association [41]. Because of the exact statistical–mechanical compensation between the enthalpic (ΔH) and entropic (ΔS) solvent–solvent (vv) contributions to the free enthalpy ΔG of association of two *neo*-pentanes in water or in urea, $\Delta H_{vv} = T\Delta S_{vv}$, these solvent–solvent terms do not contribute to the driving forces of association. Because they are part of the ΔH and $T\Delta S$ of association, they may mask in these experimentally measurable quantities the real driving forces: the barrier to association in water seems to be of entropic nature according to ΔH and $T\Delta S$ (top left panel), but is in fact of enthalpic nature according to ΔH_{uv} and $T\Delta S_{uv}$ (bottom left panel), which are the contributions that remain after removal of the exactly compensating ΔH_{vv} and $T\Delta S_{vv}$ terms. Unfortunately, the real driving terms, ΔH_{uv} and $T\Delta S_{uv}$, are only computable, not measurable.

A second feature of computer simulation results is that they may provoke experiments. For example, in early MD simulation studies of folding equilibria of polypeptides, it was found that the denatured or unfolded state of these molecules comprises much less relevant conformations than expected on the basis of simple folding models [42]. Figure 2 shows that the folding equilibrium β-heptapeptide comprises only 10^2–10^3 conformers instead of the 10^8 expected ones [43]. These simulation results provoked experimentalists to enhance their effort to find a methodology to characterize the rather small unfolded conformational state in terms of residual structure.

Simulation can only usefully replace experiments when its results are more accurate than the measured ones, which is for biomolecular systems rarely the case. However, such a rare example is shown in Figure 3, which displays the simulated and the measured free energies of binding of various biphenyl ligands to the oestrogen-binding domain of the oestrogen receptor [44]. The variation between the different experimental values (horizontal bars) is 4.2 kJ/mol larger than the average deviation of the simulated free energy of binding from the experimental free energy of binding.

Figure 3 also illustrates the use of the so-called one-step perturbation technique [45] to compute up to 10^8 [46,47] free energy differences for closely related systems from only two simulations. The atoms displayed as dotted spheres are treated as soft atoms, which can be replaced by any atom in the post-MD simulation calculation of the free energy of binding. With four atoms (H, F, Cl and Br) and nine substitution sites, this calculation yields $4^9 = 262144$ free energies of binding from just two MD simulations. Needless

Figure 2 | Number of clusters (conformers) of a β-heptapeptide at 340 K and at a pressure of 1 atm (■) and 1000 atm (●) as a function of time

In the upper panel, each point represents the total number of clusters (conformers) at the corresponding time point, and in the lower panel, the number of clusters (conformers) that make up 95 % of the trajectory sampled at the corresponding time point.

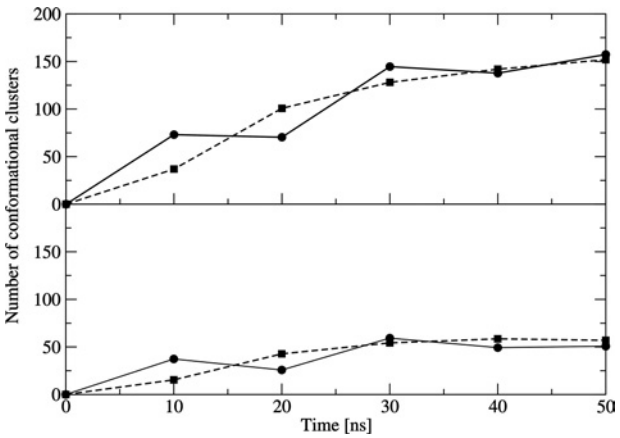
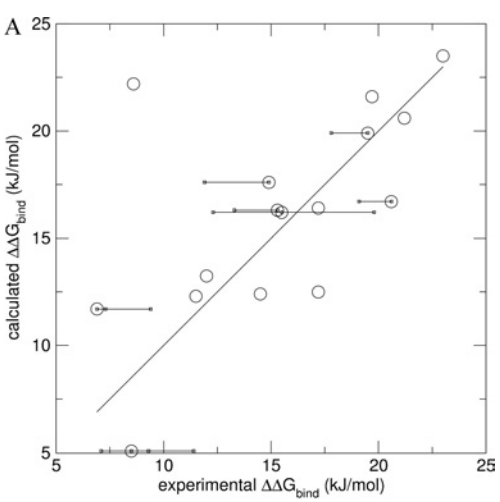
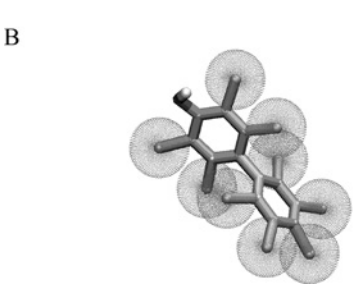


Figure 3 | Free energies of binding of biphenyl ligands to oestrogen receptor

(**A**) Experimental compared with calculated relative free energies of binding for 16 hydroxylated biphenyls. The solid diagonal line corresponds to a perfect reproduction of experimental values. Horizontal lines connect the different experimental values for a single compound. (**B**) Unphysical reference state for polychlorinated biphenyls. The nine substitution sites are shown as dotted spheres.





to say, this technique may be of use in the protection of intellectual property by patent application.

Perspectives

Present day biomolecular dynamics simulation is limited in its application by four major problems [7]: (i) the force-field problem, (ii) the search (sampling) problem, (iii) the ensemble (sampling) problem and (iv) the experimental problem. Regarding each of these issues, progress is still being made and expected. Polarizability will be introduced in standard biomolecular calculations. Force fields will be extended to cover in a consistent manner a variety of non-aqueous solvents or co-solvents. Hybrid quantum–classical [QM/MM (quantum mechanical/molecular mechanical)] simulation will benefit from improved accuracy and efficiency of the quantum part of the calculation, which currently determines its overall accuracy. In order to be able to simulate very large systems, such as proteins embedded in membranes or the ribosome or nucleosome, simplification of molecular models by averaging over atomic degrees of freedom, so-called coarse graining, is a necessity. Multiscale simulation, i.e. combining fine-grained and coarse-grained models in a consistent manner, will allow the simulation of larger systems over larger time scales without losing the possibility of recovering atomic detail when necessary [48,49], more than enough to do for the next 30 years.

Financial support by the NCCR (National Centre of Competence in Research) (Structural Biology) of the SNSF (Swiss National Science Foundation) and by the Slovenian Research Agency (ARRS) is gratefully acknowledged.

References

1 Berendsen, H.J.C. (ed.) (1976) Proceedings of the CECAM Workshop on Models for Protein Dynamics, CECAM, Orsay, France, pp. 1–304
2 van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angew. Chem. Int. Ed. Engl. **29**, 992–1023
3 Kollman, P. (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem. Rev. **93**, 2395–2417
4 Levitt, M., Hirshberg, M., Sharon, R. and Daggett, V. (1995) Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Comput. Phys. Commun. **91**, 215–231
5 Warshel, A. (1997) Computer Modeling of Chemical Reactions in Enzymes and Solutions, John Wiley and Sons, New York
6 Karplus, M. and McCammon, J.A. (2002) Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. **9**, 646–652
7 van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glattli, A., Hünenberger, P.H. et al. (2006) Biomolecular modeling: goals, problems, perspectives. Angew. Chem. Int. Ed. Engl. **45**, 4064–4092
8 Adcock, S.A. and McCammon, J.A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem. Rev. **106**, 1589–1615
9 Gee, P.J. and van Gunsteren, W.F. (2006) Numerical simulation of the effect of solvent viscosity on the motions of a β-peptide heptamer. Chem. Eur. J. **12**, 72–75
10 McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. Nature **267**, 585–590
11 van Gunsteren, W.F. and Berendsen, H.J.C. (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. **34**, 1311–1327
12 Hünenberger, P.H. and van Gunsteren, W.F. (1997) Empirical classical interaction functions for molecular simulation. in Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications (van Gunsteren, W.F., Weiner, P.K. and Wilkinson, A.J., eds), pp. 3–82, Kluwer Academic Publishers, Dordrecht
13 Lifson, S. and Warshel, A. (1968) Consistent force field for calculations of conformations vibrational spectra and enthalpies of cycloalkane and n-alkane molecules. J. Chem. Phys. **49**, 5116–5129
14 Hagler, A.T., Huler, E. and Lifson, S. (1974) Energy functions for peptides and proteins. 1. Derivation of a consistent force field including hydrogen bond from amide crystals. J. Am. Chem. Soc. **96**, 5319–5327
15 Nemethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992) Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J. Phys. Chem. **96**, 6472–6484
16 Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. **117**, 5179–5197
17 Jorgensen, W.L., Maxwell, D.S. and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. **118**, 11225–11236
18 MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S. et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B **102**, 3586–3616
19 Oostenbrink, C., Villa, A., Mark, A.E. and van Gunsteren, W.F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J. Comput. Chem. **25**, 1656–1676
20 Billeter, S.R., Webb, S.P., Iordanov, T., Agarwal, P.K. and Hammes-Schiffer, S. (2001) Hybrid approach for including electronic and nuclear quantum effects in molecular dynamics simulations of hydrogen transfer reactions in enzymes. J. Chem. Phys. **114**, 6925–6936
21 van Gunsteren, W.F. and Karplus, M. (1981) Effect of constraints, solvent and crystal environment on protein dynamics. Nature **293**, 677–678
22 van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J. and Postma, J.P.M. (1983) Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. Proc. Natl. Acad. Sci. U.S.A. **80**, 4315–4319
23 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of carbonium ion in reaction of lysozyme. J. Mol. Biol. **103**, 227–249
24 Field, M.J., Bash, P.A. and Karplus, M. (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J. Comput. Chem. **11**, 700–733
25 Hockney, R.W. and Eastwood, J.W. (1988) Computer Simulation Using Particles, 2nd edn., Institute of Physics Publishing, Bristol
26 Kastenholz, M.A. and Hünenberger, P.H. (2006) Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids. J. Chem. Phys. **124**, article no. 124106
27 Kastenholz, M.A. and Hünenberger, P.H. (2006) Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. J. Chem. Phys. **124**, article no. 224501
28 Banks, J.L., Kaminski, G.A., Zhou, R.H., Mainz, D.T., Berne, B.J. and Friesner, R.A. (1999) Parametrizing a polarizable force field from *ab initio* data. I. The fluctuating point charge model. J. Chem. Phys. **110**, 741–754
29 Ponder, J.W. and Case, D.A. (2003) Force fields for protein simulations. Adv. Protein Chem. **56**, 27–85
30 Patel, S., MacKerell, A.D. and Brooks, C.L. (2004) CHARMM fluctuating charge force field for proteins: II. Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. J. Comput. Chem. **25**, 1504–1514
31 Beveridge, D.L. and DiCapua, F.M. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu. Rev. Biophys. Biophys. Chem. **18**, 431–492
32 Hsu, S.T.D., Peter, C., van Gunsteren, W.F. and Bonvin, A.M.J.J. (2005) Entropy calculation of HIV-1 Env gp120, its receptor CD4, and their complex: an analysis of configurational entropy changes upon complexation. Biophys. J. **88**, 15–24

33 Christen, M. and van Gunsteren, W.F. (2008) On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review. J. Comput. Chem. **29**, 157–166
34 Tai, K. (2004) Conformational sampling for the impatient. Biophys. Chem. **107**, 213–220
35 Fitch, B.G., Germain, R.S., Mendell, M., Pitera, J., Pitman, M., Rayshubskiy, A., Sham, Y., Suits, F., Swope, W., Ward, T.J.C. et al. (2003) Blue matter, an application framework for molecular simulation on Blue gene. J. Parallel Distrib. Comput. **63**, 759–773
36 Shirts, M. and Pande, V.S. (2000) Computing: screen savers of the world unite. Science **290**, 1903–1904
37 Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and van Gunsteren, W.F. (1985) A protein structure from nuclear magnetic resonance data: *lac* repressor headpiece. J. Mol. Biol. **182**, 179–182
38 Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic R-factor refinement by molecular dynamics. Science **235**, 458–460
39 Yu, H., Kohl, A., Binz, H.K., Plückthun, A., Grütter, M.G. and van Gunsteren, W.F. (2006) Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. Proteins **65**, 285–295
40 Trzesniak, D. and van Gunsteren, W.F. (2006) Catalytic mechanism of cyclophilin as observed in molecular dynamics simulations: pathway prediction and reconciliation of X-ray crystallographic and NMR solution data. Protein Sci. **15**, 2544–2551
41 van Gunsteren, W.F., Geerke, D.P., Oostenbrink, C., Trzesniak, D. and van der Vegt, N.F.A. (2007) Analysis of the driving forces for biomolecular solvation and association. In Protein Folding and Drug Design: Proceedings of the International School of Physics 'Enrico Fermi', Vol. 165 (Broglia, R. A. and Tiana, G., eds), pp. 177–191, IOS Press, Amsterdam, and SIF, Bologna
42 van Gunsteren, W.F., Bürgi, R., Peter, C. and Daura, X. (2001) The key to solving the protein-folding problem lies in an accurate description of the denatured state. Angew. Chem. Int. Ed. Engl. **40**, 351–355
43 van Gunsteren, W.F. and Gattin, Z. (2007) Simulation of folding equilibria. In Foldamers: Structure, Properties and Applications (Hecht, S. and Huc, I., eds), pp. 173–192, Wiley, Weinheim
44 Oostenbrink, C. and van Gunsteren, W.F. (2004) Free energies of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation. Proteins **54**, 237–246
45 Liu, H.Y., Mark, A.E. and van Gunsteren, W.F. (1996) Estimating the relative free energy of different molecular states with respect to a single reference state. J. Phys. Chem. **100**, 9485–9494
46 Yu, H.B., Amann, M., Hansson, T., Kohler, J., Wich, G. and van Gunsteren, W.F. (2004) Effect of methylation on the stability and solvation free energy of amylose and cellulose fragments: a molecular dynamics study. Carbohydr. Res. **339**, 1697–1709
47 Oostenbrink, C. and van Gunsteren, W.F. (2005) Efficient calculation of many stacking and pairing free energies in DNA from a few molecular dynamics simulations. Chem. Eur. J. **11**, 4340–4348
48 Christen, M. and van Gunsteren, W.F. (2006) Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. J. Chem. Phys. **124**, article no. 154106
49 Berendsen, H.J.C. (2007) Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics, Cambridge University Press, Cambridge

Received 24 August 2007
doi:10.1042/BST0360011